The FHWA Travel Model Improvement Program Workshop over the Web

The Travel Model Development Series: Part I – Travel Model Estimation

presented by
Thomas Rossi
Yasasvi Popuri
Cambridge Systematics, Inc.

April 14, 2009

1

Webinar Outline

- Session 1: Introduction – October 16, 2008
- Session 2: Data Set Preparation – November 6, 2008
- Session 3: Estimation of Non-Logit Models – December 11, 2008
- Session 4: Estimation of Logit Models – February 10, 2009

2

Webinar Outline – Note Revisions! (continued)

- Session 5: Disaggregate and Aggregate Validation Procedures – March 12, 2009
- Session 6: Advanced Topics in Discrete Choice Models – April 14, 2009
- Session 7: Highway and Transit Assignment Processes – May 7, 2009

3

Webinar Outline – Note Revisions! (continued)

- Session 8: Evaluation of Model Validation Results – June 9, 2009
- Session 9: Real Life Experiences in Model Development, Webinar Wrap-Up – July 16, 2009

4

Note on Today's Session 6

Session 6: Advanced Topics in Discrete Choice Models – April 14, 2009

- This is an optional session, requested by reviewers of the original webinar outline
- More detail, more math on logit models
- No homework
- Session 5 homework will be reviewed at the beginning of *Session 7*

5

Review: The Use of Logit Models in Transportation Planning

- **Can be used to analyze any choice made by travelers with discrete alternatives**
- **Mode choice is the most common application for which logit models are used in transportation planning**
- **But there are many other choice processes for which logit models serve well**

6

Review:
The Multinomial Logit Model

$$P(1) = \frac{\exp(v_1)}{\exp(v_1) + \exp(v_2) + \cdots + \exp(v_n)}$$

Utility functions:

$V_i = B_{0i} + B_{1i} X_{1i} + B_{2i} X_{2i} + \cdots + B_{ni} X_{ni}$

where:

B_{ki} = coefficient for variable X_{ki} for alternative i

X_{ki} = variable that explains choice for alternative i

7

Modeling Individuals Disaggregately

- The outputs of the logit models are *probabilities* for all alternatives
- In aggregate models, probabilities are treated as *shares*
- In disaggregate models, probabilities can be used to *simulate* outcomes

8

Disaggregate Models

- Each person's choices are simulated individually
- Each choice depends on previously made choices

9

Disaggregate Model
Example (Home Based Work)

1. Trip production: Choose 0, 1, or 2 trips

Then, for each trip:

2. Trip distribution: Choose attraction zone
3. Mode choice: Choose auto or transit

Then, create auto and transit trip tables…

4. Perform highway and transit assignment

10

Disaggregate Model
Example (continued)

1. Trip production: MNL (3 alts.)

$U_0 = 0$

$U_1 = B_{10} + B_{11}$ (adult) + B_{12} (worker) + B_{13} (high inc.) + B_{14} (med. Inc.) + B_{15} (male)

$U_2 = B_{20} + B_{21}$ (adult) + B_{22} (worker) + B_{23} (high inc.) + B_{24} (med. Inc.) + B_{25} (male)

11

Disaggregate Model
Example (continued)

Trip production outcome for person 1:

P(0) = 0.10 P(1) = 0.20 P(0) = 0.70

Draw a random number R (0-1):

If R = 0 - 0.10, person makes 0 work trips

If R = 0.10 - 0.30, person makes 1 work trip

If R = 0.30 - 1.00, person makes 2 work trips

12

Disaggregate Model
Example (continued)

Then, for each work trip:

2. Run logit destination choice model, obtain probabilities, simulate outcome (attraction zone)
3. Run logit mode choice model, obtain probabilities, simulate outcome (mode)

After everyone has been simulated, we have a list of trips with origins, destinations, and modes.

13

Why Do This?

- Reduce aggregation error in models
- Incorporate more variables to explain travel behavior
- Get model results for population segments

14

Generic vs. Alternative Specific Variables

- Basic rule: If variable has same value for all alternatives, alternative-specific coefficients must be used AND coefficient for one alternative must be zero
- If variable has different values for different alternatives, generic specification can be used

15

Generic vs. Alternative Specific Variables: Example 1

Consider a mode choice model with 3 alts.:
Auto, transit-walk access, transit-auto access

$U_a = B_{1a}\ IVT_{ta} + B_{2a}\ (autos_a)$

$U_{tw} = B_{0tw} + B_{1tw}\ IVT_{tw} + B_{2tw}\ (autos_{tw}) + B_{3tw}\ OVT$

$U_{ta} = B_{0ta} + B_{1ta}\ IVT_{ta} + B_{2ta}\ (autos_{ta}) + B_{3ta}\ OVT$

16

Generic vs. Alternative Specific Variables: Example 1 (continued)

In the survey data set:

- $IVT_a = IVT_{tw} = IVT_{ta}$ for all observations?
  - No, therefore IVT can have a generic coefficient
    - ($B_{1a} = B_{1tw} = B_{1ta}$)
- $autos_a = autos_{tw} = autos_{ta}$ for all observations?
  - Yes, therefore IVT cannot have a generic coefficient
    - ($B_{2a} \neq B_{2tw} \neq B_{2ta}$)
  - AND, one of B_{2a}, B_{2tw}, or B_{2ta} must = 0

17

Generic vs. Alternative Specific Variables: Ease of Interpretation

If there are generic variables in the model:

- Interpreting model results easier if one alt. designated as "base alternative" for all generic variables (including the constant).
  - $B_{ka} = 0$ for all generic variables X_k

If there are only generic variables in the model:

- $B_{ka} = 0$ for all variables X_k implies that…

$$\boxed{V_a = 0}$$

18

Vehicle Availability Model
Example Estimation Results

Variable	Vehicle Availability Level 0	1	2	3	4+
Persons per household	--	--	0.1164 (2.1)	0.1164 (2.1)	0.2571 (2.1)
Workers per household	--	--	0.4915 (5.2)	1.474 (10.8)	2.139 (10.0)
Household density	--	-0.0458 (-2.9)	-0.1327 (-5.4)	-0.1717 (-4.4)	-0.2549 (-3.0)
ln(income)	--	1.130 (8.7)	2.497 (13.9)	2.995 (12.7)	3.242 (7.6)
Transit/highway accessibility	--	-1.133 (-1.7)	-2.054 (-2.8)	-2.742 (-3.3)	-2.742 (-3.3)
Persons less than vehicles	--	--	-2.870 (-8.8)	-1.017 (-5.3)	-0.5181 (1.1)
Constant	--	0.164 (0.2)	-3.761 (-4.6)	-8.229 (-8.0)	-12.87 (6.8)

ρ^2 w.r.t zero = 0.447 ρ^2 w.r.t constants = 0.302

19

Advanced Variable Specifications

- "Typical" mode choice model variables:
  - LOS: IVT, OVT (components), cost
  - Demographic (may be segmentation)
  - Zone type variables (e.g. CBD dummy, density)

20

Advanced Variable Specifications

- LOS variables:
  - Separate wait time up to X min, beyond X min
  - OVT/distance
  - % of transit IVT that is auto access
  - % of transit IVT that is local bus

21

Advanced Variable Specifications

- Demographic
  - Autos/worker, autos-workers segments (e.g. autos = 0, autos < workers, autos $\geq$ workers)
  - Consider nonlinear transformations (e.g. ln (income))
  - “Missing” income
- Combined LOS/demographic
  - Cost/income or segmented by income level

22

Size Variables

- Example: Logit destination choice (zone alts.) – number of attractions

  $V_z = \ln(Attr_z) + B_1 f(\text{travel time}) + \ldots$

- Estimated size variable

  $V_z = \ln[(\text{service emp}) + \exp(B_2)(\text{retail emp}) + \exp(B_3)(\text{other emp})] + B_1 f(\text{travel time}) + \ldots$

23

More on Interpreting Model Estimation Results

The likelihood function

$L(B) = P(c_1|B)\, P(c_2|B) \ldots P(c_n|B)$

Log-likelihood

$LL(B) = \ln P(c_1|B) + \ln P(c_2|B) + \ldots + \ln P(c_n|B)$

24

Likelihood Function
Example

Consider a binary logit model, auto vs. bus

Let V_m = a (IVT_m)

Consider a 3 trip sample:

Trip	Choice	IVT_a	IVT_B
1	A	50	30
2	A	10	20
3	B	30	40

25

Likelihood Function
Example (continued)

Choice probabilities:

1: $P(A) = 1 / [1 + \exp(-20a)]$

2: $P(A) = 1 / [1 + \exp(10a)]$

3: $P(B) = 1 / [1 + \exp(-10a)]$

26

Likelihood Function
Example (continued)

Likelihood function

$L = P(c_1|B)\ P(c_2|B)\ P(c_3|B)$

$= 1 / [1 + \exp(-20a)] \times 1 / [1 + \exp(10a)]$

$\times 1 / [1 + \exp(-10a)]$

Log-likelihood

$LL = - \ln[1 + \exp(-20a)] - \ln[1 + \exp(10a)]$

$- \ln[1 + \exp(-10a)]$

27

Use of the Likelihood Function

- Rho-squared w.r.t. zero
  - $\rho^2 = 1 - LL(B)/LL(0)$
- Rho-squared w.r.t. constants
  - $\rho^2 = 1 - LL(B)/LL(C)$

28

The Likelihood Ratio Test

1. Estimate model with all variables included. Likelihood = L1
2. Drop variables and re-estimate. Likelihood = Let log L2
3. Let LR = 2 (log L1 – log L2). LR >0.
4. LR is χ^2 distributed with k d.o.f.
5. If LR > χ^2, variables should be retained

29

Application Programming for Logit Models

- "Older" modeling software was limited in applying logit models
- Modelers often wrote stand-alone programs (FORTRAN, usually)
- Many of these legacy programs still used

30

Application Programming for Logit Models (continued)

- It is preferable to develop scripts in modeling software:
  - To input/output skims, trip tables, etc. smoothly
  - For ease in updating
  - For transparency
  - For quality control
  - For vendor support

31

Application Programming for Logit Models (continued)

- Updating older programs can be difficult
  - Commenting may be lacking
  - Input/output routines might need to be updated for newer modeling software
  - Finding the right compiler can be problematic

32

Application Programming for Logit Models (continued)

- Some hints
  - Keep estimated/calibrated parameters in a separate file
  - Keep other items that might be updated (e.g. auto operating cost) in separate file
  - Be careful with nesting coefficients
  - During debugging, have program produce interim outputs (can be commented out later)

33